DOCUMENT RESUME

ED 303 351 SE 050 335

AUTHOR Arzi, Hanna J.
TITLE From Short- to Long-Term: Studying Science Education Longitudinally.
PUB DATE 88
NOTE 39p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (61st, Lake of the Ozarks, MO, April, 1988).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150) -- Journal Articles (080)
JOURNAL CIT Studies in Science Education; v15 p17-53 1988

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Literature Reviews; *Longitudinal Studies; *Research Design; *Research Methodology; Research Problems; *Science Education
IDENTIFIERS *Science Education Research

ABSTRACT

Short-term studies often do not capture the full story of the effect that a researcher is trying to describe. If one wishes to elucidate long-term educational processes and outcomes, the only direct way of doing so is to stretch the duration of research projects and to study the same subjects over time. The limited availability of valid longitudinal data seems to be largely due to methodological problems and logistical difficulties. The focus of this article is on true longitudinal studies in science education and on other studies which aimed at studying science education over the long term. Topics discussed include: (1) long-term inferences; (2) defining and finding longitudinal studies; (3) the realities of longitudinal studies; (4) research-based knowledge in science education; and (5) longitudinal thinking. Thirty-four longitudinal studies in science education are identified and discussed in terms of focus and design features. A 124-item reference list is appended. (CW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303351

From Short- to Long-Term: Studying Science Education Longitudinally

Hanna J. Arzi

Faculty of Education
Monash University
Melbourne, Australia

Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (61st, Lake of the Ozarks, MO, April, 1988).

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Hanna J. Arzi

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE050335

Studies in Science Education, 15 (1988) 17-53

From Short- to Long-Term: Studying Science Education Longitudinally

HANNA J. ARZI
*Faculty of Education, Monash University, Melbourne, Australia**

> 'Why was the Tora likened to figs? Since the fruits of most trees, the olives, the grapes and the dates, are harvested at once, while the figs are harvested little by little. So is the Tora — today one learns a little, and tomorrow more, for it cannot be learned neither in one year nor in two.'
>
> (*Midrash Raba*, early medieval Rabbinic writings)

The term Tora in the ancient Rabbinic writings initially stood for the entire body of Jewish religious laws and learning, including both sacred literature and oral tradition: This was *the* body of knowledge worth studying. Broadly viewed from a secular perspective, the writings on the teaching and learning of the Tora pertain to the teaching and learning of any large body of knowledge of worth. Thus, applying the fig harvesting metaphor to the learning of science: Today one learns a little, and tomorrow more, for it cannot be learned neither in one year nor in two.

The obvious implication for educational researchers is that short-term studies cannot capture the full story. If one wishes to elucidate long-term educational processes and outcomes, the only direct way of doing so is to stretch the duration of research projects and to study the same subjects over time. Indeed, the desirability of longitudinal studies has been repeatedly advocated, particularly in 'Implications' and 'Recommendations' sections of articles. Yet, there is a mismatch between the consensus regarding the importance of longitudinal studies and the actual number of such studies in science education.

The present article stemmed from a concern for long-term outcomes of school science education and their gradual building-up via cumulative effects of educational experiences. This is intertwined with methodological issues: How can long-term aspects of school science education be elucidated? The limited availability of valid longitudinal data seems to be largely due to methodological problems and logistic difficulties. Hence a concern for long-term aspects of school science education subsumes a concern for how one goes about studying them.

* H.J. Arzi is now at the Centre for Science Education, P.O. Box 22685, Tel Aviv, 61226, Israel.

The issues which I have chosen to address in this article will emerge from, and be illuminated by, a review of research. The focus of my literature search was on 'true' longitudinal studies in science education, that is, studies that actually followed the same subjects over time. Complementarily, the present review also attends to studies with non-longitudinal designs — cross-sectional, retrospective, 'mixed' or 'patched-up' designs — all of which aimed at exploring science education over the long-term.

I wish to start, however, with reference to several relatively short-term studies. These studies shed lignt on the time axis beyond its origin and question the appropriateness of making long-term inferences from short-term investigations.

1. MAKING LONG-TERM INFERENCES

1.1. *The 'permanence' of measured outcomes*

When one measures outcomes beyond the completion of a given learning experience, unexpected changes may emerge over time. This was the case when Ballard (1913) set a series of experiments on the memory of children in south-east London schools. As part of these experiments which involved thousands of children, 12 year olds were asked to memorize lines from 'The Ancient Mariner' and to reproduce in writing as much as they could recall immediately after the time allotted for this task. Without being told that they would be retested, and without being given an opportunity for revision, after intervals from one to seven days the children were asked to write out the piece again. Different groups of the children were retested after different numbers of days. The average amount recalled two and three days after the memorization task was larger than the immediate recall, and at the end of the week almost 30 per cent of the children continued to retain more than they had learned initially. These findings of improved recall *after* completion of learning raised questions, since they contradicted the familiar experience of forgetting. Moreover, psychologists were bothered by the fact that they contradicted Ebbinghaus' findings of immediate forgetting (1885/1964) which had been replicated in various memory studies and had been already established as a 'classic' when Ballard reported on his investigations (1913). The conflict between Ebbinghaus' immediate forgetting and Ballard's immediate retention gains prompted a series of experiments in search of possible reasons (Luh, 1922). Apparently, the emergence of different patterns over time was due to the type of materials memorized. While Ebbinghaus studied the retention of series of unconnected nonsense syllables, Ballard investigated mainly the retention of potentially meaningful poems.

The studies cited above are not recent, nor are they concerned with meaningful learning or with science education. However, several lessons can be learned from a scrutiny of these and other long-forgotten works on memory (see, e.g., Arzi, 1985a). The point I wish to highlight now is that learning outcomes are not static. Moreover, changes occurring in learners cannot be assumed a priori to be monotonic

functions of time — increasing during interventions and staying the same or decreasing beyond it. Studies in science education, described below, support the proposition that immediate outcome measures may change over time in various ways, even within short periods. The following conclusion is that extrapolation of short-term outcome measures over the long-term can be misleading and hence is unjustified.

Findings of decrease in initially measured outcomes have been reported in both early and recent studies in science education (early studies were reviewed by Sterrett and Davis, 1954). Though the short half-lives of several outcomes may be disappointing, the phenomenon of decay is familiar, hence even results of quick and substantial losses cannot be utterly surprising. However, as will be exemplified in the subsequent paragraphs, unexpected and counter-intuitive findings have also been reported when measurements were made at more than one point in time.

The first set of examples includes investigations which detected *learning beyond instruction*, that is, studies in which the delayed outcomes were better than the immediate ones. Outcomes were found to improve 11 days after training of primary school children in classification (Chiappetta and Collette, 1975), after a 6-week retention interval in two studies with secondary students — one on conservation of volume (Rowell and Dawson, 1981) and the other on concept learning in ecology (Pella and Triezenberg, 1969), three months after instruction of the space concept in primary school (Nussbaum, 1979), and four months after a tertiary course on qualitative chemical analysis (Holcomb, 1971).

Delayed emergence of initially untraced effects is exemplified in the study of Choi and Gennaro (1987) on the use of computer simulations versus hands-on activities for enhancing understanding of the volume displacement concept in secondary school. The simulations appeared to be superior, since they led to equal performance of girls and boys, unlike the hands-on activities which favoured boys. Yet, the equivalent gender outcomes of the computer instruction were short lived: a gender gap in favour of boys emerged in a delayed posttest (45 days). Delayed emergence of effects also occurred in the study of Leonard and Lowery (1984) on questions in a tertiary biology text. Disappearance and reappearance of immediately observed effects over five weeks was reported by Kuhn and Novak (1971) in a study on advance organizers in a tertiary biology course. Dependence of change patterns on the type of the preceding treatment was found by Adejumo and Ehindero (1980) in an advance-organizer study in secondary science (three weeks), and by Padilla and Ollila (1980) in an investigation on seriation ability in primary school (three months). Different individual change patterns within a group of students exposed to the same treatment are delineated in the report of Rowell and Dawson (1980) on their study during (six weeks) and beyond (eight weeks) the instruction of the mole concept in secondary school.

A concern for 'permanence of learning' is usually accompanied by an expectation for either constancy or a decelerating decline in performance beyond the com-

pletion of a given instructional unit. These expectations have empirical support. I have chosen, however, to exemplify the dynamic nature of educational outcomes, and thus raise the issue of 'permanence' in science education, via findings which do not fully conform with such expectations. Furthermore, none of the studies cited so far extended beyond a full academic year. Hence, from a school perspective, they cannot be categorized as long-term (although in view of typical time spans of psychological experiments on memory, a retention interval of several weeks or months deserves the label 'long-term'). Despite the relatively short-term spans of these studies, multidirectional changes were observed over time. This substantiates the previously stated concern as to the validity of long-term predictions based on the extrapolation of short-term data.

1.2. *Longitudinal and quasi-longitudinal designs: an overview*

The only direct way to gain knowledge on science education over the long-term is by actually following up the same subjects from the short- to the long-term, that is, by studying science education longitudinally. However, the longer the time span of a study is, the more difficult it becomes to sustain continuous research resources and cooperation, and hence researchers are required to possess professional persistence as well as personal patience for delayed payoffs. To avoid this, short-cuts have been devised. The fact is that much of the available research-based knowledge on changes over time in science education, as well as in other areas in education, is based on replacements for true longitudinal designs.

The most popular short-cut is the *cross-sectional* research design, where data are collected at only one point in time, and longitudinal trends are inferred from differences between groups of subjects, each of which was at a different chronological stage at the point of data collection. In the other short-cut — the *retrospective* research design — longitudinal inferences are made from data collected at one point in time, on the same subjects, on both the present and previous periods (data sources may be either 'soft', such as self-reports, or 'hard' official records).

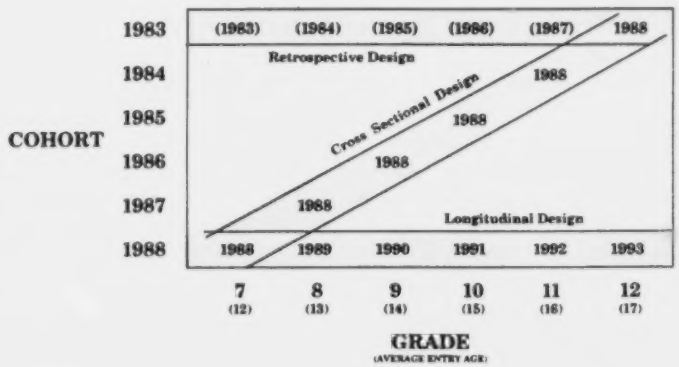
The use of the label *chronological stage* in the above definitions needs clarification. Traditionally, in biological growth studies and psychological studies on human development, change is studied across age. Accordingly, subjects are identified in terms of their age groups (or 'birth cohorts'). In educational research, particularly in studies in school contexts, the focus is not on development or maturation *per se*, but on effects of educational experiences. Hence, it is appropriate to define subjects and elapsed time in terms of school grade levels or grade-level clusters (such as lower primary school, senior high school, etc.). Although school grades are age-related, the calibration of the time-line of a study according to a school scale provides a focus on the educational context, and is conceptually different from calibration according to student age only. When the subjects of an educational study are adults, reference to elapsed time in terms of age may be of little relevance. For example, a longitudinal study on teacher change following exposure to a new science

curriculum may reveal that teacher age is a contributing variable. Yet, the time scale along which change should be monitored is years of experience with this curriculum, not teacher age.

Consequently, in the definitions of the research designs, I have used the broad notion of chronological stages, rather than ages. Similarly, this notion will be used to define the concept *cohort* as a group of subjects all of whom are at the same chronological stage at the same period, and hence share similar environmental influences at the same stages. For example, students in grades 7 to 10 in 1988 form four cohorts; grade 8 students in 1988 and students who were in grade 8 in 1959 are two distinct cohorts; likewise, teachers who embarked on a teaching career in 1988 and teachers who started teaching in 1959 can be defined as two cohorts. In studies on human development cohorts are usually defined in terms of year of birth. This is not the ultimate definition in educational research. I propose to adopt a flexible approach and to define cohorts according to the specific context of a given study. Implications of cohort differences will be illustrated later through the discussion of studies in science education. For further discussion of the cohort concept from psychological and sociological perspectives see Rosow (1978) and Ryder (1965).

The distinctions between the three basic research designs used to investigate change over time are summarized in Figure 1 via an illustration of their application to a hypothetical study on longitudinal trends in the effects of schooling during the secondary school years.

Figure 1. Illustration of alternative research designs for a hypothetical study on longitudinal trends across secondary school, launched in 1988. Cohorts are defined by year of entry to secondary school. The years in the Cohort/Grade plane denote periods of data collection; years in parentheses represent periods for which collection of data is retrospective. (The use of a Cohort/Grade plane for design description was adapted from the data matrices in Schaie, 1965.)



Both the retrospective and the cross-sectional designs would provide all the data within a single year — 1988 in the example in Figure 1, while the longitudinal data collection would linger through 1993. The longitudinal trends inferred from each of the three data sets are unlikely to be identical. The potential retrospective data sources might provide data of limited scope or of limited reliability (e.g., whatever is available in school archives, or whatever students and teachers recall). Even if the retrospective data were as rich and as reliable as the data that could be collected prospectively via a longitudinal study, both sets of data might still be different, since trends typical of 1983-1988 could be modified by 1993, owing to school reforms or a variety of contextual shifts. While in either a retrospective or a longitudinal design inferences are drawn on the basis of data collected on the *same subjects* within the *same cohort* at different points in time, in a cross-sectional study the inferences are made from comparisons between *different subjects* in *different cohorts* at the same time. The problem is that the comparability of the different cohorts cannot be checked. In fact, in the example in Figure 1, non-comparability of the cross-sectional cohorts could be expected due to selective retention of students in the upper grades of secondary school, and selective enrolment in science courses among those who stay at school. Furthermore, the curricula and environmental influences that would await the 1988 grade 7 cohort en route to grade 12 in 1993 might be quite different from the history encountered by grade 12 students prior to 1988. Consequently, loose reference to 'longitudinal data' for any sort of data which forms a basis for inferences on trends over time is unjustified: the exact research design needs to be specified.

Data from the three designs in Figure 1 are complementary, and conclusions based on analysis of the combined data would be powerful. Changes over time for the same students could be delineated, and period effects could be teased out (e.g., via comparisons between the grade 11 data sets in 1988 and in 1992). Furthermore, preliminary conclusions from the retrospective and the cross-sectional studies could assist to reformulate research questions and modify the instrumentation used in later stages of the longitudinal study. Indeed, combinations of strategies have been recommended. For example, Schaie's General Developmental Model (Schaie, 1965; Schaie and Hertzog, 1982) introduces ways for analysing combined data from replications of cross-sectional studies with different cohorts one year apart (referred to as 'cross-sectional sequences'), together with replications of longitudinal follow-ups ('longitudinal sequences').

If a choice had to be made for only one design among the three in Figure 1, on the basis of the quality of the data it could provide, regardless of the time and effort involved, it would have to be the longitudinal one (this choice is supported further in section 3 of this paper). The focus of the discussion will therefore turn to actual studies in science education which have adopted longitudinal research designs.

Readers interested in formal methodological discussions of the designs introduced above would find valuable information and ideas in the behavioural and social research literature (e.g., Goldstein, 1979; Nesselroade and Baltes, 1979; Nunnally, 1982; Plewis, 1985; Schaie and Hertzog, 1982; Wall and Williams, 1970; Wohlwill, 1973). These sources are geared primarily toward large-scale quantitative investigations, including sophisticated statistical methods aimed at deriving elaborate mathematical models for describing changes and explaining their causes. Nonetheless, the concepts underlying the discussions of the various research designs and modes of data analysis are meaningful and adaptable to both quantitative and qualitative educational research with a longitudinal orientation.

2. IN SEARCH OF LONGITUDINAL STUDIES IN SCIENCE EDUCATION

2.1. *Working definition for longitudinal studies*

There is no ultimate definition according to which one can classify unequivocally a given study as 'longitudinal'. A permissive view subsumes both retrospective and cross-sectional research designs under that label. This view is apparently shared by some researchers in science education, as can be inferred from the terminology they use to describe their methods. In view of conceptual differences between the designs, a restricted use of this label is more appropriate. According to Baltes and Nesselroade, for example: 'The one sine qua non of longitudinal research is that the entity under investigation is observed repeatedly as it exists and evolves over time.' (Baltes and Nesselroade, 1979, p.4). This statement of minimal requirements is not sufficiently clear. In particular, the notion of 'the entity under investigation' is open to interpretation. One interpretation is that exactly the same subjects are observed at all times. Schaie and Hertzog (1982) provide a statistical interpretation of the 'sameness' of the subjects that includes studies with 'sampling procedures in which a cohort is observed repeatedly by means of successive random samples from the parent population.' (Schaie and Hertzog, 1982, p.94). All in all, the succinct definition offered by Plewis (1985) describes what seems to be accepted by many as the minimal criteria for a longitudinal research design: '[A] design in which more than one member of a set of cases is measured on more than one occasion' (Plewis, 1985, p.7).

Noticeably, the minimal criteria for a longitudinal study say nothing on its time span. Accordingly, any pre-posttest study is, in effect, longitudinal. Apparently, it is taken for granted that a longitudinal study should be long, but how long? Since the present paper is concerned with *long-term* aspects of science education — particularly *school* science education, I chose to add time constraints as a minimal requirement for a longitudinal study. Specifically, the label 'longitudinal' is reserved henceforth for *studies in which information on the same subjects was collected at least twice, over a time span of at least one calendar year extending beyond the boundaries of a single school year.* The notion of 'subjects' in this definition is open to interpretation, since they can be human — students and teachers — as well as non-human entities or variables (e.g., curricular materials or achievement of a

given grade cohort). The present paper attends only to human subjects. In other words, a study would be classified as 'longitudinal' if data on the same students or teachers were actually collected at least once in each of two school (or academic) years. This working definition guided the literature search of longitudinal studies in science education.

2.2. *Longitudinal studies in science education 1963-1987*

What has been actually studied longitudinally in science education? What lessons can be learned? The list in Table 1 is believed to be representative of what has been done over the last 25 years. The choice of 1963 as the starting point for the literature search was not accidental. This was the first year of the *Journal of Research in Science Teaching* which was created in response to vigorous growth in research activities and was the first journal to be concerned solely with research in science education. Alongside articles from this journal, the list in Table 1 includes articles on longitudinal research projects published from 1963 in three other journals which are fully devoted to science education, with a major focus on research: *Science Education, European Journal of Science Education* and *Research in Science Education.*

The use of the label 'longitudinal' as a descriptor in the literature search proved to be unsatisfactory. On the one hand, as has been already noted, this label has been used loosely; on the other hand, quite a few studies which are longitudinal according to the criteria guiding the present search, did not describe themselves as such. Consequently, journals were literally searched from cover to cover. This search traced 34 longitudinal research projects. Articles emerging from the follow-up of subjects affiliated with the same entry sample were clustered in Table 1 as a single project. Different articles on the same project may refer to different dimensions studied within this project (e.g., Arzi, 1985b; Arzi, Ben-Zvi and Ganiel, 1986), or to different phases along the time-line of the study (e.g., Hadden and Johnstone, 1983a, 1983b). Articles that mention the availability of longitudinal data but do not present any were excluded. Obviously, additional aspects of some projects in Table 1 were reported through channels other than the above mentioned journals.

The studies in Table 1 will serve to illuminate issues which I have chosen to address through the rest of the present paper, though reference will be made to other studies as well. To complement the previous methodological discussion, I will start with comments on some problems associated with actual studies and reflect again on whether replacement of longitudinal by cross-sectional and retrospective designs should be endorsed (section 3). I will then attempt to provide an overview of the range of science education topics which were scanned longitudinally (section 4). This content-oriented discussion will focus on long-term concept learning and retention and revisit the phenomenon of *learning beyond instruction* which was brought to light earlier in this paper.

Table 1

Chronological Inventory of Longitudinal Studies in Science Education published during 1963-1987 in the *European Journal of Science Education, Journal of Research in Science Teaching, Research in Science Education* and *Science Education*.
(The brief notes refer to longitudinal issues only, though studies address other issues as well. The term 'wave' refers to an occasion or phase of data collection activities, and may involve the use of one or more research instruments; different waves in a given study may include identical and/or different instruments. The abbreviation exp/com stands for experimental and comparison ('control') groups.)

PROJECT	LONGITUDINAL FOCUS: *SOME FINDINGS/ CONCLUSIONS*	SOME DESIGN FEATURES
COOLEY 1964	Factors related to career decisions of males with above average ability in the process of becoming a scientist from elementary through graduate school: *Evidence for the importance of the junior high school years.*	5 years x 5 cohorts studied in parallel (starting in grades 5, 8, 11, college years 2 and 4), with 5 waves per cohort, patched up to cover 16 years (see Figure 2). (USA)
KASTRINOS 1965	Retention of biological facts taught within a preparatory college course with a principles-critical thinking approach based on Ausubel's learning theory: *High retention over 2 years.*	3 years x 1 cohort (beginning college students), 3 waves (pre- and post-course, 2 years later). (USA)
WALTERS 1965	Comparative effects of placement of a biology course in grade 9 vs 10, on learning and retention: *No differences through grade 12.*	2 sequences, 1 year apart, each of 4 years x 2 cohorts studied in parallel (grades 9-12 and 10-12), 3 waves per cohort (pre- and post-course, end of grade 12). (USA)
SMELTZ 1966	Factors affecting retention of chemistry learning in grade 11: *Theory retained over 1 year better than facts; enhanced retention after a subsequent physics course.*	2 years x 1 cohort (grades 11-12), 3 waves (pre- and post-course, end of grade 12). (USA)
GEER 1967	Long-term effects of a grade 9 extra-curricular activity of service as weather network observers: *Effects after 3 years included choice of meteorology as preferred science career, and a tendency (NS) toward better meteorology achievement.*	3 years x 1 cohort (grades 9-12) with exp/com groups, 2 waves (end of grades 9 and 12). (USA)

DAVISON & GEISERT 1968	Comparison of sequencing high school chemistry (CHEMStudy) before or after biology (BSCS): *No differences in achievement in either course.*	2 years x 1 cohort (grades 9-10) with 2 groups (chemistry before/after), several waves each year. (USA)
SCOTT 1970,1973	Changes in cognitive style during a 3-year inquiry training in upper-elementary grades (phase 1), and retention of effects through high school (phase 2): *Change in the direction of the training persisted through the training; 1 of 6 measured effects persisted 6 years beyond training.*	Phase 1: 3 years x 1 cohort (grades 5-7), 3 waves (beginning of grade 5, end of 6 and 7); Phase 2: 5 years x 1 cohort (grades 7-12), 2 waves (end of grades 7 and 12). Comparison groups were either followed up or matched ad hoc. (USA)
MACKAY 1971	Student change during 2 years of studying physics via new (PSSC-based) and traditional courses: *Mixed effects, such as net decline in enjoyment of physics vs improvement in physics-related cognitive processes; similar outcomes across new and old courses.*	2 years x 2 cohorts 1 year apart (grades 11-12) with 5 groups (new/old courses), 3 waves per cohort (beginning of grade 11, end of 11 and 12). (Australia)
MOON 1971	Teacher change during the first year of teaching a new elementary school programme (SCIS) after inservice training (3 weeks): *Initially observed change in verbal behaviour in the desired direction (indirect approach and high-order questions) dissipated over the year.*	1 year x 1 cohort (elementary school teachers) with exp/com groups, 5 waves (starting at the school year prior to the inservice summer course). (USA)
ALLEN 1972, 1973a,b	Evaluation of consecutive units within a new elementary school programme (SCIS) compared with traditional textbook teaching: *Each of the 3 SCIS units taught in grades 1-3 produced superior cognitive but not affective outcomes.*	2+ years x 1 cohort (grades 1-3) with exp/com groups, 3 waves (end of grades 1,2 and 3). (USA)
JOHNSON 1972	An integrated 2-year chemistry-physics course team-taught by chemistry and physics teachers, compared with consecutive separate courses (CHEMStudy and PSSC): *Evidence for superior outcomes of the integrated course, some effects could be detected only at the end of the course.*	2 years x 1 cohort (grades 11-12) with 2 groups (integrated/separate courses), 3 waves (beginning of grade 11, end of 11 and 12). (USA)

JUNGWIRTH & DREYFUS 1974	Development of process-oriented behaviour in preservice biology and agriculture teachers: *Significant gains were made over a 2-year programme, yet the outcomes did not reach the level of experienced exemplary teachers.*	2 years x 2 cohorts 1 year apart (preservice high school teachers), 2 waves per cohort (beginning and end of a 2-year programme). (Israel)
TOMERA 1974	Retention of processes of observation and comparison following instruction (4 weeks) in grade 7: *Substantial retention over 15 months.*	1 + years x 1 cohort (grades7-8), 4 waves (3 in grade 7). (USA)
LASHIER & NIEFT 1975	Teacher change during the first year of teaching a new junior high school programme (ISCS) in grade 7, after inservice training (5 weeks): *Significant change in science classroom activities in the desired direction (more student participation).*	1 year x 1 cohort (junior high school teachers) with exp/com groups, 2 waves (starting at the end of the school year prior to the inservice summer course). (USA)
MAYER, DISINGER & WHITE 1975	Evaluation of a 2-year inservice earth-science programme for junior high school teachers via a study of teachers during and beyond the programme: *Evidence for significant cumulative effects includes persistent change in science classroom activities in the desired direction (more inquiry oriented) and improved student understanding.*	3 years x 1 cohort (junior high school teachers), 6 waves (starting at the end of the school year prior to the inservice programme). (USA)
ORGREN & DORAN 1975; ORGREN 1977	Teacher change following mandated adoption of a new high school earth science curriculum (New York Regents), compared with teachers who had opted to adopt it earlier: *Change during the first year of mandated adoption continued over the next 3 years in the desired direction (more inquiry-oriented teaching).*	5 years x 2 cohorts studied in parallel (teachers who had opted to teach the course and teachers mandated to do so later), 5 waves per cohort (starting a year before the course became mandatory). (USA)
TAMIR & JUNGWIRTH 1975	Evaluation of student achievement and attitudes during a 4-year high school biology programme (Israel BSCS Adaptation Project): *Findings include the gradual building up of substantial gains in inquiry skills.*	4 years x 3 cohorts 1 year apart (grades 9-12), with exp/com groups, several waves over the years. (Israel)

SHELDON 1978	Implementation of new programmes following curriculum awareness conferences for elementary school teachers and administrators (2 weeks): *The number of conference participants implementing new programmes immediately (more administrators than teachers) did not change much 3 years later.*	3 years x 1 cohort (elementary teachers and administrators), 2 waves (1 and 4 years after the conferences). (USA)
DORAN & NGOI 1979	Concept learning and retention in a 2-year hierarchically-designed instruction on the particle nature of matter in elementary school grades 5-6: *Retention of basic concepts over 1 year was enhanced by review in conjunction with subsequent teaching of advanced concepts.*	2 years x 3 cohorts 1 year apart (grades 5-6), 4 waves per cohort (2 in each year). (USA)
PETERSON 1979	Change in sensory-motor curiosity in a science-enriched environment from childhood (ages 6-12) to adolescence (ages 12-18): *No change in amount of expressed curiosity.*	6 years x 1 mixed age sample (ages 6-12 to 12-18), 2 waves (6 years apart). (USA)
POWER 1979, 1981	Change in student attitudes to science during the transition between primary and secondary schools: *The majority of students enjoyed science throughout the transition period.*	2 years x 1 cohort (grades 7-8), 4 waves (beginning and end of grades 7 and 8). (Australia)
BOTTOMLEY & ORMEROD 1981	Factors affecting interest in, preference for, and choice of, science subjects, over the first 2 years of secondary school: *Several science activities in middle school correlated significantly with choices made after 2 years in secondary school.*	2 years x 1 cohort (years 1-2 in secondary school), 3 waves (beginning of year 1, end of 1 and ?). (UK)
FISHER, WILLIAMS & ROTH 1981	Effect of frequent computer-assisted self-evaluation on retention of a tertiary-level genetics course (1 term), compared with conventional testing: *The computer-assisted evaluation resulted in significantly better performance immediately after the course, and better — though NS — 2 years later.*	2 + years x 1 cohort (college students) with exp/com groups, 3 waves (pre- and post-course, 2 years later). (USA)

WRIGHT 1981 (1978)	Retention of effects of instruction in processes of cue attendance and hypothesis generation in grade 9 (1 individualized session) on open exploration behaviour: *Some significant immediate outcomes were retained over 15 months, better results for instruction in cue attendance.*	1+ years x 1 cohort (grades 9-10) with exp/com groups, 2 waves (post-grade 9 instruction and 15 months later). (USA)
MITCHELL & KELLINGTON 1982	Learning and retention of the particulate structure of matter in an introductory secondary school science course: *Forgetting over a year beyond instruction attributed to insufficient use of acquired concepts in subsequent instruction.*	1 + years x 1 cohort (years 1-3 in secondary school), 2 waves (post-instruction in year 1 and beginning of year 3). (UK)
WESTERBACK 1982	Change in attitude to, and anxiety about, teaching science, in preservice elementary school teachers: *Positive change during a year of science content courses persisted through student teaching in the subsequent year.*	2 years x 1 cohort (preservice elementary school teachers), 4 waves (starting at the beginning of content courses). (USA)
HADDEN & JOHNSTONE 1983a,b	Change of attitudes to science through the first 2 years of secondary school: *Attitudes declined in science more than in other subjects, change appeared to have been influenced by ethos of school science departments.*	2+ years x 1 cohort (end of primary to end of year 2 in secondary school), 3 waves (1 each year). (UK)
AULT, NOVAK & GOWIN 1984	*(As part of a methodological paper, partial data from a longitudinal project are presented:)* Case studies of change in children's molecule concepts through primary school with or without grade 1 instruction: *Concept meanings grasped in primary grades may affect levels of understanding observed years later.*	5+ years x 1 cohort (grades 2-7) with exp/com groups, 2 waves. (USA)
HANDLEY & MORSE 1984	Gender differences in change of self-concept and gender-role perceptions in science from grade 7 to 8, and their relationships to achievement and attitude: *Gender differences emerged in grade 8, more changes observed with attitude than with achievement.*	1 year x 1 cohort (grades 7-8), 2 waves (end of grades 7 and 8). (USA)

STRAWITZ 1984	Retention of ability to control variables following training in grade 6 (4 weeks) with teacher-directed vs open-ended approaches: *Effects were retained 1 year beyond treatment, better results via the teacher-directed instruction.*	1 year x 1 cohort (grades 6-7), with 2 treatment groups (a comparison group added later), 3 waves (pre- and post-treatment, 1 year later). (USA)
WILSON & WILSON 1984	Development of formal thinking during 2 years of pretertiary science courses: *Students developed significantly in terms of their Piagetian levels.*	2 years x 1 cohort (grades 11-12), 2 waves (beginning of grade 11, end of 12). (Papua New Guinea)
ARZI 1985b; ARZI, BEN-ZVI & GANIEL 1986	Learning and retention of physical science in sequential science courses in grades 8-10: *Evidence for retroactive facilitation of retention by curriculum continuity, and for retained knowledge which cannot be readily retrieved yet is functionally available for further learning.*	3 years x 2 cohorts 1 year apart (grades 8-10) with 2 groups (more/less chemistry instruction), several waves per cohort each year. (Israel)
GENNARO, HEREID & OSTLUND 1986	Exploration of long-term effects of family science courses (5 weekends) for grades 6-8 children and their parents: *Mixed findings on influences on learning behaviour and parent-child relationships 2 years beyond the course.*	2 years x mixed age group of children (grades 6-8 to 8-10) and their parents, 2 waves (starting at the family courses). (USA)
KELLY 1986	Change in attitudes to science during the first 3 years of secondary school, and effects of an intervention aimed at enhancing girls' attitudes to, and achievement in, science: *Overall attitude decline with slight intervention effects for both genders and large variations among schools; attitude stability within individuals.*	2+ years x 1 cohort (years 1-3 in secondary school) with exp/com groups, 2 waves (starting at the beginning of year 1). (UK)

3. REALITIES OF LONGITUDINAL STUDIES

3.1. *Longitudinal research congruity*

Despite the expectation that the working definition introduced above would enable me to classify studies unequivocally, the decision to include or exclude studies in or from Table 1 was not always clear. For example, Wright (1981) reached conclusions on student long-term retention of process skills by using a Posttest-Only

Control-Group Design (according to Campbell and Stanley's terminology, 1963). Having data on only one point in time, this study cannot be classified as longitudinal. Yet, a previous article (Wright, 1978) provides relevant data on the same cohort 15 months earlier. Thus, the combination of the 1978 and the 1981 papers is, in effect, one project which subscribes to the definition of a longitudinal study.

The question whether or not the combination of Wright's papers makes a 'legitimate' longitudinal study is not just a technical matter. It points to the distinction between collection of longitudinal data, analysis of such data, and drawing conclusions on longitudinal patterns. Consequently, this raises the question of conceptual and methodological congruity from the phases of formulating research questions and designing a study through the phase of reporting on it. This issue can be illustrated further through the evaluation of a science curriculum (SCIS) in elementary schools in Honolulu, which was reported by Allen in a series of articles (including Allen, 1972, 1973a, 1973b). The reports refer to a follow-up of the same students within experimental and control groups ('SCIS' versus 'non-SCIS') in grades 1 to 3. The data collected at each year were analysed and reported separately, with no discussion across two points in time. The link between reports is preserved only via the introductory comments which state that this is part of a longitudinal study, and inform of the current rate of attrition of the entry sample (from 300 students in grade 1 to 213 in grade 2 and 176 in grade 3). Since attrition is frequently selective (as will be discussed in section 3.2.), the comparability of the separate sets of analyses is not guaranteed, and readers who wish to construct their own interpretation of change over time are unable to do so satisfactorily. The question is, therefore, why should a researcher go to the trouble of tracing students in different schools over several years, when — eventually — data at each of the points in time are treated as discrete entities? In the case of Allen's SCIS study, it is possible that the researcher planned to analyse the data longitudinally at a later stage and may have done so in a publication which was not retrieved in my search. Whatever the case may be, it served to illustrate incongruity within the process leading from the design of, to the report on, a given study.

Longitudinal studies are valuable, but a study is more than its design. Noticeably, the definition adopted in this paper identifies a study in terms of the apparent features of its design: same subjects, at least two data collections (waves), at least one year. The underlying assumption that led me to propose such a definition was that a researcher would embark on a long-term project with a longitudinal design only if she or he had *longitudinal goals,* and that the later phases of analysis and reporting would be matched accordingly. In other words, a longitudinal design — similarly to any research design — is expected to match the research goals and to be matched by subsequent research operations. Some studies in Table 1 do not conform fully with this assumption. Indeed, Baltes and Nesselroade (1979) found it necessary to include the notion of goals in the definition they offer for the context of developmen-

tal psychology: '*Longitudinal methodology involves repeated time-ordered observation of an individual or individuals with the goal of identifying processes and causes of intraindividual change and interindividual patterns of intraindividual change in behavioral development.*'(Baltes and Nesselroade, 1979, p.7). It is unclear, however, how goals should be assessed: on the basis of the overt goal statements of researchers, or by inferences from the data they actually collected and how they analysed them?

Obviously, it is easier to give a technical description of a design of a study than to define its goals. Since I believe that a design which subscribes to the technical criteria of being longitudinal implies longitudinal goals, I propose to use the extent of conceptual-methodological congruity between the design and the research goals, as well as between the design and subsequent research stages, as an important measure of research quality. While all the studies in Table 1 are longitudinal per definition (with a few border line cases), they vary in quality. The evaluation of quality of each of them is likely to vary among readers with different personal interpretations.

3.2. *Threats to longitudinal studies*

A frequent design among the projects in Table 1 and among longitudinal studies in general consists of a single cohort followed by two or more waves. This is actually a modified version of the One-Group Pretest-Posttest Pre-Experimental Design (O X O) which was introduced by Campbell and Stanley as a ' "bad example" . . . worth doing where nothing better can be done' (1963, p.177). Following these harsh words, Campbell and Stanley provided a constructive discussion of factors jeopardizing the validity of this and of less criticized designs which form bases for studies in Table 1. This and related discussions available in the literature provide valuable guidelines for any research project, since the list of 'threats' is essentially the same for either short- or long-term studies. Yet, some threatening factors become more prominent as the time span of a study increases. Among these factors are (using again the terminology of Campbell and Stanley, 1963) *maturation, history, mortality* and *selection*. The factor known as *testing* which refers to the mutual effects of successive administrations of instruments may, in fact, mellow due to longer time intervals, although cumulative effects due to many measurements over time may be manifested eventually.

As has been already noted, the foci of methodological discussions vary according to the research thrust and context. This is evident in the different ways in which maturation and history effects are treated in the realm of research on human development and in the realm of educational research — particularly in school contexts. For science educators, 'pre' and 'post' usually refer to a quasi-experimental intervention or to points in time along an existing sequence of school science courses the effects of which they wish to describe or explain. Other effects during the pre-post

interval — maturation and history — are therefore regarded as threats to valid conclusions. For developmental psychologists, 'pre' and 'post' refer to points in the life span of subjects in a given age cohort, and maturation is the focus of interest — not a threat. According to the latter research perspective, effects of special programmes in school science might be considered as non-normative influences jeopardizing the validity of conclusions on what are thought to be 'true developmental influences and processes'. For science educators studying science concept learning from a constructivist perspective, both effects of schooling and maturation should be of interest — not threats, and gaining insights into their interplay over the long-term should be a major research challenge. None of the studies in Table 1 was committed fully to this thrust.

Among the items in the list of threats to educational research in general and to longitudinal research in particular, the most bothering problem is probably the continuous and practically unavoidable loss of subjects from the entry sample. This is frequently referred to as *attrition*, sometimes as 'erosion', as well as by the less pleasant terms 'wastage' and 'mortality' (literal mortality is a real issue in life-span developmental studies). In high-status longitudinal studies with extensive funding and large staffs, attrition can be reduced substantially (see, for example, the activities conducted to retain subjects in the National Longitudinal Study, Davis and Levinsohn, 1979). Yet, such studies are not typical of science education. Major reasons for attrition in longitudinal studies in school settings include mobility of students between schools — especially in transition from one school level to another (e.g., primary to secondary), and the need to sustain cooperation with different staff along several years. Idiosyncratic reasons may also interfere, as can be exemplified by the closing of a school due to political unrest during the study of Wilson and Wilson (1984) in Papua New Guinea. Loss of entire schools from entry samples, however, is not restricted to such dramatic circumstances (e.g., Bottomley and Ormerod, 1979; Hadden and Johnstone, 1983b; Wright, 1981).

Attrition is a 'logistic headache' that interferes with the neat progress of elegantly designed studies and necessitates large entry samples and tedious book-keeping. More important, it is frequently selective, and hence gradually biases the sample compared with its entry characteristics. In designs with control or comparison groups, differential attrition with differential selection may occur across groups. There is evidence that students lost from school studies tend to have lower academic achievement than those who survive longitudinal investigations. This also seems to be the case when students are not lost irreversibly from the entry sample, but are absent in one measurement occasion and reappear in the next. Selection in such cases may be because those who are frequently absent are likely to include a large proportion of less motivated or less healthy students with lower achievement. Consequently, despite the fact that a longitudinal study intends to follow the same subjects, data sets at different points in time may not be fully comparable. To secure

the validity of longitudinal conclusions, they have to be based on analysis of *longitudinally matched data,* that is, data pertaining to exactly the same subjects at all points in time. Since the subjects for whom full longitudinal data are available may comprise a select group relative to the entry sample, the generalizability of conclusions drawn from their data may be limited. In other words, the external validity of internally valid longitudinal conclusions should be checked.

The following example of the book-keeping and matching needed due to attrition and concomitant selection is based on my experience in a study on various aspects of retention in sequential physical science learning from junior- to senior-high school (Arzi, 1983). Reasons for attrition in this study were similar to those mentioned above. The drop in numbers was also affected by a decision not to search for all students but to concentrate the follow-up efforts mainly in schools where the subjects of the entry sample had a reasonably large representation. At each measuring occasion, all the students in the relevant classes were tested, whether or not they had been part of the entry sample. The additional data thus gained were later used to cross check the external validity of the longitudinal results. All in all, following an entry sample of 3167 students at the beginning of grade 8, the sample at the beginning of grade 9 consisted of 2512 students including 1866 from the entry sample, and a year later in grade 10 — 2289 students with 1176 representatives of the entry sample. Comparisons between entry data on the dropouts and those who survived the longitudinal investigation showed that the attrition changed the entry sample both quantitatively and qualitatively: the follow-up samples were smaller and academically better. Since the study was multidimensional, the necessary longitudinal matching was done separately for each of the dimensions. Thus, inferences on retention from grade 8 through grade 10 were based on five different data sets which were merged and matched, eventually leading to a file with complete 3-year data on 506 students (Arzi et al., 1986). Another research question demanded different data, the matching of which resulted in a file on 1816 students (Arzi et al., 1985). Some of the follow-up and matching strategies exemplified through this study are similar to those used in the 8-year ETS Study of Academic Prediction and Growth (see the description of the ETS study by Hilton (1979), and the empirical comparisons made by Hilton and Patrick (1970) between matched and unmatched data, as well as between longitudinal and cross-sectional data from the same study).

While the detailed example above related to a largely quantitative study, attrition also threatens qualitative studies, since it reduces the richness of the data across time, and thus may limit the extent to which valid interpretations could be proposed.

Research articles often provide only scanty descriptions of research methods, probably due, in part, to lack of sufficient journal space. In longitudinal studies, information on attrition and its causes, checks for selection and subsequent matching (if any), are not petty details. Articles which do not present such information, or

present it unclearly, do not enable readers to appreciate their conclusions, and thus make only a limited contribution. Furthermore, if one believes that more and better longitudinal studies can contribute to the advancement of science education, then sharing of concrete research experiences is imperative.

3.3. *Alternatives to longitudinal studies?*

In view of the difficulties associated with the actual conduct of longitudinal studies, the question of alternatives is raised again. To complement the previous general discussion of other designs used for making longitudinal inferences (section 1.2.), concrete examples from the science education research literature will be examined.

3.3.1. *Cross-sectional designs: an alternative?*

'As important as cross-sectional studies of age groups are in their own right, they provide only gross approximations to true longitudinal trends . . . Few people these days would argue that it is sensible to construe purely cross-sectional data as definitive of longitudinal trends' (Nunnally, 1982, p.134). In science education, however inferences on change over the long-term (or lack of it) have been made on the basis of cross-sectional studies more often than on the basis of longitudinal studies. The major problem is *cohort effects* which preclude full comparability of different cross-sectional cohorts. It is therefore bothering when cross-sectional research reports provide insufficient information on the extent to which their cohorts are comparable, or ignore this issue altogether. Furthermore, inferences on changes *within individuals* are drawn from cross-sectional data, although such data inform only on differences *between groups*.

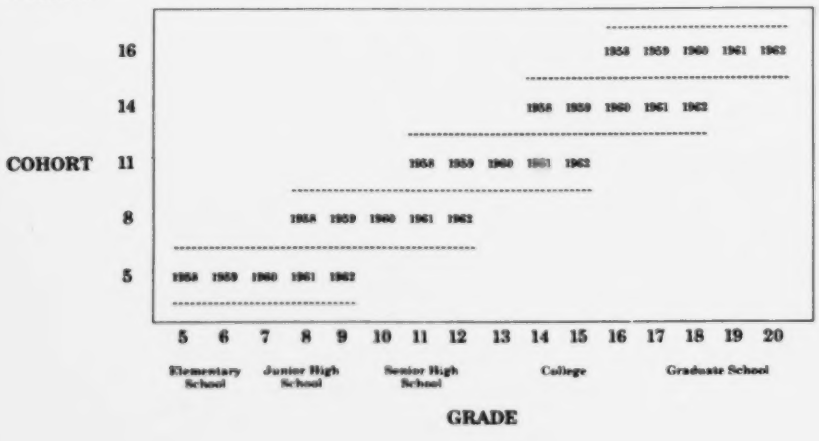
The report of Rakow and Walker (1985) on the status of Hispanic American students in science exemplifies careful conclusions from a cross-sectional study. Comparisons of science achievement and experience of students in three cohorts — ages 9, 13 and 17 — showed that the gap between white and Hispanic students (in favour of whites) decreased across age. Yet, Rakow and Walker (1985) point out that the high rate of Hispanic students who drop out from high school resulted in a selective population of Hispanic students in the upper school grades, and consequently the younger and the older age cohorts could not be compared. A longitudinal study following 9 year olds through age 17 would have lost the less able (or the less high achieving) Hispanic students along the way, but the data could have been longitudinally matched and valid conclusions could have been made as to whether the gap between whites and a selective group of Hispanic students has indeed changed through school.

Apart from drop-out from school, comparability of cohorts in school science education is often affected by differential uptake of science courses. For example, in a cross-sectional study on attitudes to laboratory work, Hofstein, Ben-Zvi and Samuel (1976) observed that attitudes in grade 11 were more positive than either

in grade 10 or grade 12. They suggested that since only one third of the grade 10 cohort was likely to choose science courses in grades 11 and 12, the attitude difference between grades 10 and 11 reflected selection while the grades 11 and 12 data reflected a longitudinal trend. Incomparability of cross-sectional cohorts enrolled in different courses was also encountered in the study of Simpson and Oliver (1985) in science classes from grade 6 through grade 10. Findings of this study showed a general attitude decline across grades, with the exception of grade 9. The related discussion discloses the information that the grade 9 cohort included primarily advanced students who elected to study physical science, while the rest of the student population was heterogeneous. Cohort incomparability is not mentioned again in the discussion of motivation to science achievement across grades, probably because the grade 9 data fitted in the apparent continuous decline from grade 6 to grade 10.

The examples above dealt with comparability of cohorts in school beyond compulsory education and beyond compulsory science courses. One would expect better comparability between primary school grade cohorts. This was the assumption underlying the cross-sectional study of Nussbaum (1979) on children's conceptions of the Earth as a cosmic body. Yet, the trend across grades was not as clear as one would have hoped. Indeed, the one cohort which particularly interfered with a neat trend and blurred instructional effects was found to be different in terms of students' IQ scores and families' socio-economic status. Nussbaum's cautious conclusions on the existence of longitudinal trends in children's conceptions of the Earth, based on his cross-sectional data on Israeli students, were supported by cross-sectional studies in Nepal (Mali and Howe, 1979) and in the USA (Sneider and Pulos, 1983). Yet, only a longitudinal study would elucidate directly the *intra-individual* changes

Figure 2. A research design with longitudinal and cross-sectional features, illustrated by the Career Development of Scientists Study according to its description by Cooley, 1964. Cohorts are defined by grade upon entry into the study. The years in the Cohort/Grade plane denote periods of data collection.



in children's science concepts across time, and how this could be affected by instruction.

Although cross-sectional studies can provide only approximations to longitudinal trends, they can result in valuable data *if* they are undertaken carefully. Questions arising from puzzling quasi-longitudinal findings of a cross-sectional study deserve a true longitudinal follow-up.

3.3.2. *Cross-sequential (patched-up) designs: an alternative?*

A compromise between a longitudinal and a cross-sectional design is exemplified in Figure 2 by the design chosen for the Career Development of Scientists Study (Cooley, 1964). This was described by Cooley as 'an overlapping longitudinal study'. Such designs are also known as 'cross-sequential' (Schaie, 1965) or 'patched-up' (Campbell and Stanley, 1963).

As can be seen in Figure 2, five true longitudinal sequences enabled Cooley to cover a period of 16 years within five years. The overlapping data for different cohorts provided opportunities for checking on the comparability of the longitudinal sequences. Irrespective of the elegant design of this study, some of its conclusions probably have only limited applicability nowadays, in view of the fact that it was concerned with the career development of high-ability males only, and in view of the generational shifts over the three decades that have elapsed since its inception.

Although patched-up designs, such as Cooley's (1964), are considered legitimate by many, they have been also criticized: '[E]fforts have been made to piece together different segments of longitudinal data . . . into an approximation of what would have been obtained had the same subjects been followed from beginning to end. This effort leads one into a quagmire of assumptions and dubious statistical practices . . . [A]ny effort to piece together a precise developmental function from a hodgepodge of partially longitudinal strands is doomed to failure.' (Nunnally, 1982, p.134). Although 'a precise developmental function' is not a high priority goal in science education, this criticism is conceptually important. In any case, cross-sequential designs are superior to the simple cross-sectional designs which abound in science education research, and are worth considering when a full longitudinal study is impossible or impractical.

3.3.3. *Retrospective designs: an alternative?*

Retrospective designs are not a well-defined entity. They cover a range of strategies, from data collection via unstructured interviews in which subjects are asked to share their reminiscences, to systematic search searches of school records or archives. Retrospective designs have been frequently used in attempts to relate success in tertiary education and career patterns to science in secondary school, as well as in the study of various other topics (e.g., Deboer, 1984; Diamond, St. John, Cleary and Librero, 1987; Fensham, 1974; James, 1972; Neujar and Hansen, 1970; Reitz, 1973; Tamir and Amir, 1981; Willson, 1980). In some cases, retrospective

studies are only semantically different from true longitudinal studies. For example, retrospective collection of student science grades and prospective (i.e., longitudinal) collection of such grades at the time they are produced (waiting patiently from one school term to the next) would result in identical data.

School files may sometimes provide more versatile data than grades and demographic information, as was the case in the study of Wynn and Bledsoe (1967) on changes in science interest during the high school years. This investigation was done in a school where, as a normal practice, the interests of all grade 9 students were measured every year. Wynn and Bledsoe measured the science interest of students in grades 11 and 12 and matched each student's data with her or his grade 9 interest scores and other data from the school files. Thus,despite the retrospective rather than longitudinal collection of data, the resulting data were *longitudinal in effect*. Consequently, the findings (showing considerable time stability of science interest) are more valid than findings based on data from cross-sectional comparisons of non-comparable cohorts. The advantage of an analogous true longitudinal study would be that it would not have to be restricted by retrospectively available data. Instead, it could attempt to create a rich data base in accordance with a priori specified research questions.

Although, as a rule, retrospective studies cannot replace longitudinal studies, the following examples show that they can play valuable complementary roles.

The first example is the study of Bottomley and Ormerod (1981) on the influence of middle school science activities on the science interest of secondary school students. The entry sample consisted of students in three secondary schools fed by 52 different middle schools. Information on science activities in middle school was collected retrospectively by student reports at the beginning of secondary school. This was correlated with the data on student interest which were collected longitudinally over the next two years. Since the correlated variables were measured at different points in time, the results of the correlational analysis could also suggest causality.

The use of retrospective data within longitudinal projects can also be exemplified via a study on the development of science teachers' knowledge of science (Arzi, White and Fensham, 1987). The entry sample consisted of prospective teachers doing their post-graduate preservice training course. As part of the first data collection, each participant was asked to try to assess retrospectively the sources which had made contributions to the accumulation of their knowledge on certain science concepts. This was followed by a longitudinal study on the effects of teaching experiences on teacher knowledge. Since the teachers' prior knowledge sources spanned over many years and across both formal and non-formal contexts (school, university, media, personal experiences), such information could not have been collected directly via a formal longitudinal study.

4. LONGITUDINAL RESEARCH-BASED KNOWLEDGE IN SCIENCE EDUCATION

4.1. *Coverage of topics in longitudinal studies 1963-1987*

The only common denominator of longitudinal studies is the general frame of their research design. Thematically, longitudinal studies in science education are a diverse group, as can be seen from the notes in Table 1. These studies followed students and teachers, and include evaluations of curricular and extra-curricular projects, studies on grade placement and sequencing of subject matter, monitoring of effects within or beyond preservice and inservice teacher training courses, studies on process and content learning and instruction, investigations of changes in student affective measures (attitudes, interest, curiosity), science choices, career patterns and gender issues. Although the scope of themes that have been studied longitudinally over the last 25 years is impressive, the number of studies per theme is small, and the few studies within a theme may be different from each other in terms of the research questions, contexts and methodologies. Consequently, despite interesting results from individual studies, it is difficult to make reference to cumulative longitudinal knowledge about each theme. I will attempt to make a few comments on findings and missing findings on a few topics, before dwelling at more length on one theme: concept learning.

The value of longitudinal studies to science education can be exemplified by insights they provided into science attitudes. The studies of Bottomley and Ormerod (1981), Kelly (1986) and Hadden and Johnstone (1983) in English and Scottish schools, confirmed findings of many cross-sectional studies across countries on a trend of decline through the secondary school years. Yet, these studies also pointed to exceptions and uncovered variations among schools which allude to specific school-related influences on change patterns. Kelly's (1986) interesting finding on the stability of attitude structures *within* individuals, in parallel with the overall decline, could not have been elicited in a cross-sectional study. Power's (1981) observation of no attitude decline in the transition of Australian students from primary to secondary school questions a common belief that traces the origin of the secondary school attitude decline to this transition.

In view of the on-going debate on the relationships between attitudes and achievement (Schibeci, 1984), it is surprising that the present search did not retrieve longitudinal studies on this issue. Since longitudinal studies can inform on the time order of attitudes and achievements, they are more appropriate for making causal inferences from attitude-achievement correlations than studies which measure the correlated variables at the same point. In other words: if x (e.g., science attitude) is significantly correlated with y (e.g., achievement), and if evidence was visible for x before it was available for y, then x can be a cause of y, but not vice versa. Thus, although longitudinal studies cannot prove causality (neither can any correlational study), they may help to rule out causal directions which are incompatible with the time-order of the correlated variables, and hence are appropriate for

the study of attitude-achievement relationships. For formal discussions of causal modelling in longitudinal studies see Plewis (1985) and Rogosa (1979).

An important theme with a relatively thin coverage in Table 1 is curriculum evaluation. One might have expected more, in view of the repeated endorsement of long-term evaluation (e.g., Cronbach, 1963; Shulman and Tamir, 1973; Tyler, 1949). The available longitudinal evaluations include both studies which monitored effects within long programmes, and studies which assessed long-term effects beyond the completion of either short or long programmes. Several projects attempted to evaluate curriculum innovations and related teacher inservice courses via longitudinal studies of teacher change. Out of four such studies in Table 1, the one with primary teachers reported dissipation of effects over the one year follow-up (Moon, 1971), while the studies with secondary teachers reported persistent effects over one to four years (Lashier and Nieft, 1975; Mayer, Disinger and White, 1975; Orgren, 1977). Despite similarities, these studies are not fully comparable, hence it is impossible to tease out the factors affecting persistent teacher change. This is an issue worth further longitudinal exploration.

The theme of grade placement and sequencing of subject matter is sometimes subsumed under the broad heading of curriculum evaluation. There is little valid longitudinal data to support or question related decisions at the school level. Grade placement and sequencing of science subject matter continue to be determined largely by tradition, or by value judgements of subject specialists and curriculum developers. For example, the extent to which the introductory physical science course IPS provided an appropriate basis for CHEMStudy was evaluated positively on the basis of analysis of logical sequential relationships (D'Amboise and Searles, 1974); likewise, a survey of opinions of teachers and professors led to the conclusion that the high school chemistry course, CBA, should be a good basis for tertiary-level chemistry (Montean, Cope and Williams, 1963). The present literature search did not find empirical evidence for the latter conclusions, yet retrieved three longitudinal studies in American high schools which raised questions on the American tradition of placing biology in grade 10, followed by chemistry in grade 11 and physics in grade 12 (Davison and Geisert, 1968; Johnson, 1972; Walters, 1965). Each of these studies set out to explore specific questions in the context of a single school. For example: Should chemistry (CHEMStudy) precede biology (BSCS)? (Davison and Geisert, 1968); Should chemistry (CHEMStudy) and physics (PSSC) be taught integratively? (Johnson, 1972). Underlying these seemingly curriculum-specific and context-specific questions are broad issues about the cumulative effects of science courses on student long-term growth. One major aspect — cumulative effects on long-term science concept learning — will be the focus of the last part of the discussion on longitudinal research-based knowledge in science education.

4.2. *Long-term concept learning: Reflections and tentative conclusions from an insufficient research base*

4.2.1. *Evidence of 'conceptual stability despite instruction'*

Despite the fact that science educators have been always interested in content achievement, and despite vigorous research activities over more than a decade on children's concept understanding and discrepancies between 'alternative' and 'scientific' views of the world (e.g., Driver, Guesne and Tiberghien, 1985; Osborne and Freyberg, 1985; West and Pines, 1985; White, 1988), there is little valid longitudinal data on long-term learning and retention in science (see Table 1). Consequently, a large part of current knowledge on science content learning, including concept learning, can be validly applied to short-term situations only. In the following discussion I would like to question, in particular, the generalizability of the repeatedly cited proposition 'student alternative conceptions resist change *despite* instruction'.

The labelling of student pre-instructional ideas as 'change/instruction resistant' is usually based on either cross-sectional studies of student conceptions across age, or on relatively short-term intervention studies. The unjustifiability of the extrapolation of short-term data over the long-term, and problems with longitudinal inferences from cross-sectional data, have been examined previously in this paper. It cannot be ignored, however, that stability of student ideas in science across age groups has been repeatedly observed in careful cross-sectional studies in different countries. A problem that continues to exist in the face of these replications is that cross-sectional data can give no information on stability of concepts *within* individuals. Making gross approximations on change (or lack of it) from differences between groups of different students is contrary to the ethos among researchers on student ideas who believe primarily in in-depth probes at the individual level.

Two recent longitudinal studies probed student ideas individually. Novak and his group followed the same cohort of students throughout their schooling, from grade 1 to 12 (Novak, 1986), and Engel Clough, Driver and Wood-Robinson (1987) probed changes within secondary school students over a 2-year interval — from age 12 to 14 in one cohort and 14 to 16 in another. Both studies present evidence for conceptual stability within individuals. For example, Engel Clough et al. report that more than half the responses given on tasks on different concepts — both 'correct' and 'incorrect' — remained unchanged over two years. Yet, the results of both studies also show that patterns of change are more complex than what one might infer simplistically from cross-sectional data averages. This is illustrated visually in Novak's study through concept maps describing the molecule concepts of the same students in different grades (see concept maps from this study in Ault, Novak and Gowin, 1984, Figures 5-8). Moreover, Novak's study also provides evidence that quality instruction can have lasting effects. The findings show that children who were exposed to an audio-tutorial unit on the particulate structure of matter at grade 1 had improved understanding of the model, including fewer misconceptions, through grade 12, relative to children who had not received such instruction.

A major limitation of cross-sectional studies on student change patterns concerns their ability to tease out instructional effects. In order to make valid inferences on the value of science concept instruction, detailed information is needed on the presence or absence of deliberate attempts to deal with student alternative ideas. Such data are difficult to obtain in cross-sectional studies, since the only way to collect them is retrospectively. Furthermore, cohort incomparability is likely to blur instructional effects, masking subtle effects altogether.

The caution needed in generating and generalizing conclusions on instructional effects can be illustrated via many studies — short-and long-term, cross-sectional and longitudinal. For example, Andersson and Kärrqvist (1983), in their study of Swedish students' understanding of light, criticize the way in which optics was presented to grade 8 students. Hence, the findings on poor understanding of light among grade 9 students, despite previous instruction, can be interpreted as a case of poor teaching rather than a clear case of conceptual tenacity. In another fine study, Shipstone (1984) detected improved understanding of electricity among students in the upper years of secondary school, compared with younger cohorts, on which he comments: 'It is difficult to be certain what proportion . . . results from teaching rather than from self-selection of courses by those students who are more able at physics' (p.189). Another issue which was difficult to ascertain in this study is the extent to which the popularity of a certain faulty current model among 15 year olds was due to some idiosyncracy of this cohort, or — as Shipstone suspects — due to the first exposure of these students to voltage without sufficient differentiation between voltage and current. Issues like this could be resolved via longitudinal studies, provided the studies followed students' ideas in parallel with instruction and beyond it. Longitudinal studies making only some reference to instruction have limited contribution to the understanding of the interplay between instruction and concept learning over the long-term.

4.2.2. *Longitudinal facilitation of concept learning*

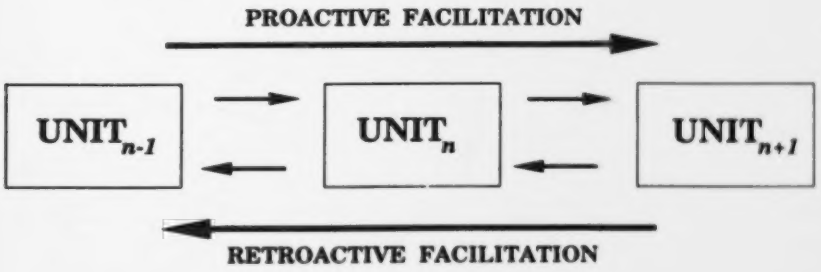
The long-term payoffs of continuous and coordinated learning experiences, as opposed to discrete instructional units, was highlighted by findings from a longitudinal study on effects of interrelations between courses on acquisition and long-term retention of physical science concepts (Arzi, 1983). In one part of this study, student knowledge of the particulate structure of matter was followed two years beyond the instruction of the topic in grade 7. It was found to be significantly better in a group of students who continued to study related physical science topics in grade 8, relative to students whose subsequent science courses did not include such topics. These long-term learning outcomes were not due to mere rehearsal, since the content of the grade 7 course was used but not retaught in subsequent science units (Arzi et al., 1985). This phenomenon of facilitation of previous learning by new related materials emerged again with other concepts investigated in this

study (Arzi, 1985b; Arzi et al., 1986). These results, obtained in Israeli secondary schools, support inferences from longitudinal data obtained in Scottish secondary schools (Mitchell and Kellington, 1982), where low retention of concepts associated with the particulate structure of matter was attributed to disuse of concepts following their initial learning. These results are also in accord with the findings of a longitudinal study with American students in the upper grades of elementary school (Doran and Ngoi, 1979) which show the benefits of a 2-year sequential instruction of hierarchically organized concepts of the particle model.

Potential retroactive effects of subsequent learning on the consolidation and retention of related prior knowledge could be expected on the basis of Ausubel's learning theory (Ausubel, 1963; Ausubel, Novak and Hanesian, 1978). When a student is exposed to new materials which are related to antecedent learning, and if the student indeed relates the new materials to her or his prior knowldege, then the interaction between new and prior knowledge during the assimilation process mobilizes previously learned concepts in the student's cognitive structure, thus increasing their availability for future use. Furthermore, clarification and consolidation of the prior knowledge may occur via *progressive differentiation* of concepts, in conjunction with *integrative reconciliation* of new and old meanings. In other words: once the encounter of a student with formal instruction of a given topic is over, learning is not necessarily over. The more often the learner encounters related concepts, and the more links she or he establishes between new and old concepts, the easier it is to retrieve items for further use and the more consolidated the whole cognitive structure becomes. Apparently, in the above longitudinal studies on learning of the particulate structure of matter, different groups of students were exposed to different instructional experiences beyond their first encounter with the model, hence the variations in the extents to which their initial learning was retroactively facilitated.

The notion of *retroactive facilitation* was first suggested by Ausubel and supported via two short-term experiments on learning and retention of passages dealing with Christianity, Buddhism and Zen Buddhism (Ausubel, Robbins and Blake, 1957; Ausubel, Stager and Gaite, 1968). The experiments and their findings provoked debate (reviewed by Ausubel et al., 1978, and by E.D. Gagné, 1978), and the idea of retroactive facilitation has never gained popularity (unlike the popularity of the counter idea of retroactive interference which is apparently typical of rote rather than meaningful learning). The research report of Arzi et al. (1985) was the first attempt to attend to the educational implications of this idea for long-term school science learning. Clearly, more longitudinal investigations are needed for better understanding of the phenomenon and its application to concept learning. I believe, however, that in view of the currently available evidence, the idea of retroactive facilitation of learning, together with the more familiar and the more studied idea of proactive facilitation effects of prior knowledge, provides a framework for designing curriculum and instruction toward the enhancement of long-term con-

Figure 3. Longitudinal model for efficient sequential learning.



cept learning (Figure 3). A constructivist approach to curriculum development in science, as presented by Driver and Oldham (1986), fits into this longitudinal model.

Instruction of concepts should proceed more often via longitudinally coordinated units where students are provided with opportunities to encounter the same concepts in different contexts beyond their initial formal instruction. If each unit of instruction looks back in an attempt to build on student ideas *and* to consolidate previously acquired knowledge, and if at the same time each unit looks forward in an attempt to help students to construct a basis for further learning, then — perhaps — student alternative ideas in science would prove to be less resistant to change than the current evidence suggests.

One of the bothering issues concerning the longitudinal model for efficient sequential learning (Figure 3) is the distinction between potential and actual facilitation of learning. This was evident, for example, in the 2-year investigation of Davison and Geisert (1968) in search of the assumed payoffs of sequencing chemistry (CHEMStudy) before biology (BSCS), rather than the other way round. Although chemistry concepts taught in the chemistry course were included in the biology course (not vice versa), the texts were not coordinated, and the teachers in the study deliberately avoided making explicit references to antecedent or subsequent materials in order to preserve research rigour. The results, showing that placing chemistry before or after biology made no difference, suggest that students did not actualize the potential for facilitation of learning, and probably assimilated the contents of the two courses into largely separate chemistry and biology schemata.

The currently available longitudinal data in science education which led to the model of efficient sequential learning in Figure 3 is not enough. More is needed to substantiate concept-specific suggestions as to how to coordinate concept instruction longitudinally, and as to how to increase the extent to which the potential of longitudinally coordinated science concept instruction is realized in terms of student growth of concept understanding over time. Suggestions, such as science-related training in metalearning (or metacognition), including use of concept maps as a

learning device (Novak and Gowin, 1984) to assist students to interlink concepts across instructional units and across science disciplines, may prove to be helpful. Yet, there is not sufficient science-related longitudinal data on this and other strategies.

Furthermore, current learning theories do not offer much guidance on the long-term. R.M. Gagné (1985) admits that 'little is known about what specifically happens to learned materials during the time it is stored in long-term memory' (p.73); in his discussion on enhancing of retention and transfer he adds: 'It may be hoped that much more work will be forthcoming about this question than is currently known' (p.255). A similar view is expressed in a recent review of cognitive conceptions of learning (Shuell, 1986): 'The more important forms of human learning that interest most cognitive scientists and educators involve what is fundamentally a long-term process involving weeks, months, and even years. The phases that we go through as we engage in long-term learning are unknown at present' (p.429).

Apparently, Ausubel's learning theory (Ausubel et al., 1978) attends to long-term issues more than currently prevailing theories in cognitive psychology (Arzi and West, 1986). Though Ausubel's discussion of long-term issues is not sufficiently detailed and parts of it need further empirical substantiation, the theory offers guidelines which are applicable to school science concept learning. The above mentioned ideas of long-term effects of prior knowledge, retroactive facilitation of learning and retention, progressive differentiation and integrative reconciliation, provide a theoretical frame for the design of longitudinal studies on science concept learning. Accumulation of a critical mass of valid longitudinal data will, it is to be hoped, ground an improved model for long-term science learning.

4.2.3. *Time for learning during and beyond instruction*

The idea of processes that consolidate previous learning and counteract post-instructional forgetting and regression was supported in the preceding discussion by longitudinal data. Additional evidence can be found in short-term studies presented at the beginning of this paper (the experiments of Ballard (1913) and several science retention studies) in which the expected decline in outcomes beyond instruction was counterbalanced — at least for a while. The same phenomenon was also repeatedly observed during the first days immediately following completion of interventions, as reported by Mayer and his colleagues (e.g., Kwon and Mayer, 1985) who used intensive time-series designs in their research programme on science concept learning.

A scrutiny of the contexts of each of the cases in which enhancement of learning outcomes was evident beyond the formal completion of an instructional unit on a given topic, shows that this could not be always attributed to direct or indirect effects of further instruction. Apparently, although changes in students' science knowledge are likely to occur due to curriculum continuity (Arzi et al., 1985, 1986), they do not necessarily depend on it. The environment — both natural and human-

made — is a rich source of potential stimuli. Similarly to retroactive facilitation of previous learning by subsequent school materials, interactions with non-school stimuli activate concepts in a student's cognitive structure thus counteracting forgetting and sometimes leading to knowledge consolidation.

One could expect, however, that students with the same school experiences would vary in their exposure to, awareness of, and extent of interaction with, non-school science-related stimuli. This provides a plausible explanation for gender differences in science. Indeed, data from large scale cross-sectional studies in Canada (Erickson and Erickson, 1984) and in the UK (Johnson and Murphy, 1984) support the attribution of the superior science achievement of boys — at least in part — to their superior accumulation of knowledge via exposure to more science-related experiences. I propose to extend the interpretation of the experiential hypothesis on gender differences by considering also the differential accumulation of post-instructional science-related experiences and their retroactive effects on consolidation of school acquired knowledge. This and other facets of the experiential hypothesis still await elucidation.

Longitudinal changes in student understanding of school-taught concepts which cannot be attributed to definite school experiences are particularly difficult to assess. They may evolve through student self-stimulated reflections (e.g., due to contradictions between prior ideas and newly taught knowledge), they may also be initiated via idiosyncratic experiences long after instruction. I will provide anecdote-type evidence about Suzanne — a science teacher participating in a longitudinal study on teacher knowledge and its development (Arzi et al., 1987). After 12 years of schooling, a BSc(Hon), completion of a full year post-graduate preservice course for science teachers, and a year of secondary school science teaching, Suzanne thought that last week (i.e., a week before the research interview) bits of her knowledge 'clicked' meaningfully in regard to what the conservation of matter was all about in terms of conservation of atoms. This occurred during a visit with her father (who has no formal science background beyond compulsory education) when their fireside chat drifted toward a discussion of the newly built hearth and 'where does everything disappear when it burns inside'.

In view of Driver's (1985) review of findings (non-longitudinal) on difficulties encountered by students in the understanding of the conservation of matter, I believe that Suzanne's long-term difficulties may be typical of many. Yet, my randon content analysis of introductory physical science and chemistry texts shows that prior understanding of this idea, particularly its manifestation in terms of atoms, is sometimes taken for granted. Whatever the exact reasons for Suzanne's lagging understanding of the conservation of matter, her case shows that 'elementary' concepts continue to develop over long periods.

The conclusion about the slow and complex process of science concept learning has been previously suggested in this paper on the basis of non-anecdotal longitudinal data, particularly from studies on the learning of the particulate structure of mat-

ter (the only topic for which I retrieved more than a single longitudinal study). The longitudinal data are further complemented by data from non-longitudinal studies. Although analogies between student concept learning and the history of science cannot be carried too far, I would still emphasize that generations of scientists struggled to establish what we now regard as 'elementary' concepts. In regard to the particle model, it was not until the famous Karlsruhe Congress in 1860 (half a century after Dalton had proposed his ideas) that chemists resolved their differences in respect of atoms versus molecules (Ihde, 1964).

In view of the emerging consensus for the constructivist perspective of learning, one would expect that concept learning is not a simple linear process in which the application of a dose of instruction implies an immediately observable increment of conceptual change. Construction of knowledge takes time, and it is reasonable to expect stages of 'incubation' during which concept meanings ripen without visible progress. Hence, concept learning is unlikely to follow skill acquisition curves which typically show continuous improvement along a training programme, usually with particularly fast improvements at the beginning of training (Lane, 1986). The kinetics of concept learning are likely to include lag-phases, and — as many studies show — full understanding (whatever that is) is frequently not achieved before the time allotted for instruction is over. Yet, low achievement after a given instructional unit may mask the establishment of foundations which can be functional in further construction of knowledge.

What we can currently generalize across science concepts is that the processes involved with learning them are long and complex. The process of studying science concept learning is similarly long and complex.

5. 'THINKING LONGITUDINALLY' — CLOSING COMMENTS

> 'The first object of any act of learning, over and beyond the pleasure that it may give, is that it should serve us in the future. Learning should not only take us somewhere, it should allow us to go further more easily.'
>
> (Bruner, 1960, p.17)

Schools are thought to prepare students for the future. The long-term outcomes of schooling are expected to be produced gradually, by cumulative effects of educational experiences. The cumulative and long-term nature of education, however, does not seem to have a high priority on the agenda of educational researches, including researchers in science education. The result is insufficient valid research-based knowledge on the *long-term* processes involved with school science teaching and learning and their outcomes.

The intention in writing this paper was to increase the extent of 'longitudinal thinking' in science education through the provision of some conceptual and methodological guidelines. The paper is therefore issue-driven, not method-driven. The focus on studies with a longitudinal research design reflects the view that they provide the most appropriate way for tackling long-term issues. Although all the

longitudinal studies referred to in this paper were concerned with changes in students or teachers, the longitudinal methodology also applies to the study of important questions concerning changes in non-human entities and variables over time (such as changes during the adaptation of curricular materials or trends in attitudes and achievement of a given age group over the years).

Clearly, longitudinal studies are not a panacea for science education. The fact that a study is longitudinal does not guarantee its quality or its significance. There are many important issues which can be elucidated satisfactorily by studies which do not follow the same subjects over time, and there are long-term issues which can benefit from well designed and carefully conducted non-longitudinal studies.

The fact remains, however, that there are too few longitudinal studies, and that we know too little of some issues where the preferred research design should be longitudinal. It appears that the avoidance of a methodology led to the neglect of research areas. In the review of science education research over the last decade, White and Tisher (1986) made such an observation in regard to research on Ausubel's theory: 'Most research concentrated on a relatively minor part of Ausubel's theory . . . Advance organizers are a simple idea, readily amenable to group experiments, while Ausubel's other ideas about assimilation of knowledge are more complex and may require detailed longitudinal studies of individuals, which so far researchers appear to have found daunting.'(p.883). The criticism that research is sometimes guided by easy methods rather than important questions, probably applies to more than one area. For example, it is surprising that despite the abundance of Piagetian studies in the science education literature, none of the longitudinal studies retrieved in the present search in major journals sets out to investigate Piaget's important concepts of cognitive development.

It is true that longitudinal studies are not easy enterprises, but they are probably more feasible than is commonly thought — a conviction regarding their importance for generating knowledge definitely helps. Novak's 12-year study (1986) is an exemplary case, showing how professional leadership, coupled with conviction and persistence, tied together research efforts of several 'generations' of doctoral students into a non-fragmented piece of work.

Large scale and highly funded longitudinal studies have their place. Science education could benefit no less from budget-affordable in-depth longitudinal studies in well defined contexts. Replications of such studies in different contexts would help to generate the much needed knowledge on important issues in science education. This is a slow and long process. Yet, as has been suggested by the fig-harvesting metaphor at the beginning of this paper, knowledge of worth can be accumulated only little by little.

Acknowledgement. I wish to thank Glenn L. Rowley and Richard T. White of Monash University for reviewing the manuscript.

REFERENCES

ADEJUMO, D. and EHIDERO, S. (1980). Facilitating learning of science-oriented textual materials in a developing country: Study in the use of organizers. *Science Education, 64,* 397-403.

ALLEN, L.R. (1972). An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the Interaction Unit of the Science Curriculum Improvement Study elementary science program. *Journal of Research in Science Teaching, 9,* 167-173.

ALLEN, L.R. (1973a). An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the Systems and Subsystems Unit of the Science Curriculum Improvement Study elementary science program. *Journal of Research in Science Teaching, 10,* 125-134.

ALLEN, L.R. (1973b). An examination of the ability of third grade children from the Science Curriculum Improvement Study to identify experimental variables and to recognize change. *Science Education, 57,* 135-151.

ANDERSSON, B. and KÄRRQVIST, C. (1983). How Swedish pupils, aged 12-15 years, understand light and its properties. *European Journal of Science Education, 5,* 387-402.

ARZI, H.J. (1983). *Proactive and retroactive effects of courses in sequential physical science learning: A longitudinal evaluation.* Unpublished-doctoral dissertation, Weizmann Institute of Science, Rehovot, Israel.

ARZI, H.J. (1985a). Educational reflections provoked by the centennial of Ebbinghaus' nonsense syllables (1885-1985). In *Collected Papers of the 1985 Annual Conference of the Australian Association for Research in Education* (pp.354-357). Hobart, Tasmania: AARE.

ARZI, H.J. (1985b). The long-term kinetics of conceptual development: The case of + = − and contextual differentiation. *Research in Science Education, 15,* 112-121.

ARZI, H.J., BEN-ZVI, R. and GANIEL, U. (1985). Proactive and retroactive facilitation of long-term retention by curriculum continuity. *American Educational Research Journal, 22,* 369-388.

ARZI, H.J., BEN-ZVI, R. and GANIEL, U. (1986). Forgetting versus savings: The many facets of long-term retention. *Science Education, 70,* 171-188.

ARZI, H.J. and WEST, L.H.T. (1986). *Ausubel revisited: New insights from a seemingly exhausted source.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

ARZI, H.J., WHITE, R.T. and FENSHAM, P.J. (1987). *Teachers' knowledge of science: An account of a longitudinal study in progress.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

AULT, C.R., Jr., NOVAK, J.D. and GOWIN, D.B. (1984). Constructing Vee maps for clinical interviews on molecule concepts. *Science Education, 68,* 441-462.

AUSUBEL, D.P. (1963). *The Psychology of Meaningful Verbal Learning.* New York: Grune and Stratton.

AUSUBEL, D.P., NOVAK, J.D. and HANESIAN, H. (1978). *Educational Psychology — A Cognitive View* (2nd ed.). New York: Werbel and Peck.

AUSUBEL, D.P., ROBBINS, L.C. and BLAKE, E., Jr. (1957). Retroactive inhibition and facilitation in the learning of school materials. *Journal of Educational Psychology, 48,* 334-343.

AUSUBEL, D.P., STAGER, M. and GAITE, A.J.H. (1968). Retroactive facilitation in meaningful verbal learning. *Journal of Educational Psychology, 59,* 250-225.

BALLARD, P.B. (1913). Obliviscence and reminiscence. *British Journal of Psychology, Monograph Supplements, 1* (2, Whole No.2).

BALTES, P.B. and NESSELROADE, J.R. (1979). History and rationale of longitudinal research. In J.R. Nesselroade and P.B. Baltes (Eds.), *Longitudinal Research in the Study of Behavior and Development* (pp.1-39). New York: Academic Press.

BOTTOMLEY, J. and ORMEROD, M.B. (1981). Stability and lability in science interest from middle schools to the age of science choices (14 +). *European Journal of Science Education, 3,* 329-338.

BRUNER, J.S. (1960). *The Process of Education.* Cambridge, MA: Harvard University Press.

CAMPBELL, D.T. and STANLEY, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of Research on Teaching* (pp.171-246). Chicago: Rand McNally.

CHIAPPETTA, E.L. and COLLETTE, A.T. (1975) The effectiveness of verbal label training in aiding second grade pupils to transfer their classification skill. *Journal of Research in Science Teaching, 12,* 183-191.

CHOI, B.S. and GENNARO, E. (1987). The effectiveness of using computer simulated experiments on junior high students' understanding of the volume displacement concept. *Journal of Research in Science Teaching, 24,* 539-552.

COOLEY, W.W. (1964). The potential scientist pool (PSP). *Journal of Research in Science Teaching, 2,* 24-32.

CRONBACH, L.J. (1963). Course improvement through evaluation. *Teachers College Record, 64,* 672-683.

D'AMBOISE, A. and SEARLES, W.E. (1974). Introductory Physical Science (I.P.S.) a useful background for Chem.Study students. *Journal of Research in Science Teaching, 11,* 341-350.

DAVIS, J.A. and LEVINSOHN, J.R. (1979). The NLS Study of the High School Class of 1972: A resource for educational and human development researchers. *New Directions for Testing and Measurement, 2,* 69-94.

DAVISON, J. and GEISERT, P. (1968). Can and should chemistry precede biology? *Science Education, 52,* 364-368.

DEBOER, G.E. (1984). A study of gender effects in the science and mathematics course-taking behavior of a group of students who graduated from college in the late 1970s. *Journal of Research in Science Teaching, 21,* 95-103.

DIAMOND, J., St. JOHN, M., CLEARY, B. and LIBRERO, D. (1987). The Exploratorium's Explainer Program: The long-term impacts on teenagers of teaching science to the public. *Science Education, 71,* 643-656.

DORAN, R.L. and NGOI, M.K. (1979). Retention and transfer of selected science concepts in elementary school students. *Journal of Research in Science Teaching, 16,* 211-216.

DRIVER, R. (1985). Beyond appearances: The conservation of matter under physical and chemical transformations. In R. Driver, E. Guesne and A. Tiberghien (Eds.), *Children's Ideas in Science* (pp.145-169). Milton Keynes, UK: Open University Press..

DRIVER, R., GUESNE, E. and TIBERGHIEN, A. (Eds.). (1985). *Children's Ideas in Science.* Milton Keynes, UK: Open University Press.

DRIVER, R. and OLDHAM, V. (1986). A constructivist approach to curriculum development. *Studies in Science Education, 13,* 105-122.

EBBINGHAUS, H. (1964). *Memory: A Contribution to Experimental Psychology* (H.A. Ruger and C.E. Bussenius, Trans.). New York: Dover. (Original work published 1885).

ENGEL CLOUGH, E., DRIVER, R. and WOOD-ROBINSON, C. (1987). How do children's scientific ideas change over time? *School Science Review, 69,* 255-267.

ERICKSON, G.L. and ERICKSON, L.J. (1984). Females and science achievement: Evidence, explanations, and implications. *Science Education, 68,* 63-89.

FENSHAM, P.J. (1974). Long-term effects of science education at school. *Research in Science Education, 4,* 11-20.

FISHER, K., WILLIAMS, S. and ROTH, J. (1981). Qualitative and quantitative differences in learning associated with multiple-choice testing. *Journal of Research in Science Teaching, 18,* 449-464.

GAGNÉ, E.D. (1978). Long-term retention of information following learning from prose. *Review of Educational Research, 48,* 629-665.

GAGNÉ, R.M. (1985). *The Conditions of Learning and Theory of Instruction* (4th ed.). New York: Holt, Rinehart and Winston.

GEER, I.W. (1967). A follow-up study of the effects of service as mesoscale weather network observers. *Journal of Research in Science Teaching, 5,* 350-354.

GENNARO, E.D., HEREID, N. and OSTLUND, K. (1986). A study of the latent effects of family learning courses in science. *Journal of Research in Science Teaching, 23,* 771-781.

GOLDSTEIN, H. (1979). *The Design and Analysis of Longitudinal Studies.* London: Academic Press.

HADDEN, R.A. and JOHNSTONE, A.H. (1983a). Secondary school pupils attitudes to science: The year of erosion. *European Journal of Science Education, 5,* 309-318.

HADDEN, R.A. and JOHNSTONE, A.H. (1983b). Secondary school pupils attitudes to science: The year of decision. *European Journal of Science Education, 5,* 429-438.

HANDLEY, H.M. and MORSE, L.A. (1984). Two-year study relating adolescents' self-concept and gender role perceptions to achievement and attitudes toward science. *Journal of Research in Science Teaching, 21,* 599-607.

HILTON, T.L. (1979). ETS Study of Academic Prediction and Growth. *New Directions for Testing and Measurement, 2,* 27-44.

HILTON, T.L. and PATRICK, C. (1970). Cross-sectional versus longitudinal data: An empirical comparison of mean differences in academic growth. *Journal of Educational Measurement, 7,* 15-24.

HOFSTEIN, A., BEN-ZVI, R. and SAMUEL, D. (1976). The measurement of the interest in, and attitudes to, laboratory work amongst Israeli high school chemistry students. *Science Education, 60,* 401-411.

HOLCOMB, C.M. (1971). The effect of degrees of direction in the qualitative analysis laboratory on retention of learning. *Journal of Research in Science Teaching, 8,* 165-169.

IHDE, A.J. (1964). *The Development of Modern Chemistry.* New York: Harper and Row.

JAMES, R.J. (1972). Traits associated with the initial and persistent interest in the study of college science. *Journal of Research in Science Teaching, 9,* 231-234.

JOHNSON, G. (1972). An integrated two year chemistry-physics course compared with consecutively taught separate courses. *Science Education, 56,* 143-154.

JOHNSON, S. and MURPHY, P. (1984). The underachievement of girls in physics: Towards explanations. *European Journal of Science Education, 6,* 399-409.

JUNGWIRTH, E. and DREYFUS, A. (1974). Biology-teachers' on-the-spot decisions (Some problems in preservice teacher education). *Science Education, 58,* 205-214.

KASTRINOS, W. (1965). A study of the retention of biological facts by high school biology students. *Science Education, 49,* 487-491.

KELLY, A. (1986). The development of girls' and boys' attitudes to science: A longitudinal study. *European Journal of Science Education, 8,* 399-412.

KUHN, D.J. and NOVAK, J.D. (1971). A study of cognitive subsumption in the life sciences. *Science Education, 55,* 309-320.

KWON, J.S. and MAYER, V.J. (1985). Identification and description of the momentum effect in studies of learning: An abstract science concept. *Journal of Research in Science Teaching, 22,* 253-259.

LANE, N.E. (1986). *Skill acquisition curves and military training.* IDA Paper P-1945. Alexandria, V.A: Institute for Defence Analyses.

LASHIER, W.S., Jr. and NIEFT, J.W. (1975). The effect of an individualized, self-paced science program on selected teacher, classroom, and student variables — SCIS Level One. *Journal of Research in Science Teaching, 12,* 359-369.

LEONARD, W.H. and LOWERY, L.F. (1984). The effects of question types in textual reading upon retention of biology concepts. *Journal of Research in Science Teaching, 21,* 377-384.

LUH, C.W. (1922). The conditions of retention. *Psychological Monographs, 31* (3, Whole No.142).

MACKAY, L.D. (1971). Changes in physics students in three states over a two year period. *Research in Science Education (Research 1971), 1,* 109-118.

MALI, G.B. and HOWE, A. (1979). Development of Earth and Gravity concepts among Nepali children. *Science Education, 63,* 685-691.

MAYER, V.J., DISINGER, J.F. and WHITE, A.L. (1975). Evaluation of an inservice program for earth science teachers. *Science Education, 59,* 145-153.

MITCHELL, A.C. and KELLINGTON, S.H. (1982). Learning difficulties associated with the particulate theory of matter in the Scottish Integrated Science course. *European Journal of Science Education, 4,* 429-440.

MONTEAN, J.J., COPE, R.C. and WILLIMAS, R. (1963). An evaluation of CBA chemistry for high school students. *Science Education, 47,* 35-43.

MOON, T.C. (1971). A study of verbal behavior patterns in primary grade classrooms during science activities. *Journal of Research in Science Teaching, 8,* 171-177.

NESSELROADE, J.R. and BALTES, P.B. (Eds.). (1979). *Longitudinal Research in the Study of Behavior and Development.* New York: Academic Press.

NEUJAR, J. and HANSEN, R. (1970). The prediction of careers in science from high school data. *Journal of Research in Science Teaching, 7,* 391-394.

NOVAK, J.D. (1986). *A 12 year longitudinal study of science concept learning*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

NOVAK, J.D. and GOWIN, D.B. (1984). *Learning How to Learn*. Cambridge, UK: Cambridge University Press.

NUNNALLY, J.C. (1982). The study of human change: Measurement, research strategies and methods of analysis. In B.B. Wolman (Ed.), *Handbook of Developmental Psychology* (pp.133-148). Englewood Cliffs, NJ: Prentice Hall.

NUSSBAUM, J. (1979). The effect of SCIS's 'Relativity' unit on the child's conception of space. *Journal of Research in Science Teaching, 16*, 45-51.

NUSSBAUM, J. (1979). Children's conceptions of the Earth as a cosmic body: A cross age study. *Science Education, 63*, 83-93.

ORGREN, J.R. (1977). The long-term effects of mandated curriculum adoption on teaching behavior. *Journal of Research in Science Teaching, 14*, 419-425.

ORGREN, J.R. and DORAN, R.L. (1975). The effects of adopting a revised New York State Regents Earth Science Syllabus on selected teacher and student characteristics. *Journal of Research in Science Teaching, 12*, 15-24.

OSBORNE, R. and FREYBERG, P. (Eds.). (1985). *Learning in Science: The Implications of Children's Science*. Auckland, New Zealand: Heinemann.

PADILLA, M.J. and OLLILA, L. (1980). Effect of small-group teaching on acquisition and transfer of non visual seriation ability. *Science Education, 64*, 357-366.

PELLA, M.O. and TRIEZENBERG, H.J. (1969). Three levels of abstraction of the concept of equilibrium and its use as an advance organizer. *Journal of Research in Science Teaching, 6*, 11-21.

PETERSON, R.W. (1979). Changes in curiosity behavior from childhood to adolescence. *Journal of Research in Science Teaching, 16*, 185-192.

PLEWIS, I. (1985). *Analysing Change*. Chichester, UK: Wiley.

POWER, C. (1979). Science achievement and attitudes at primary and secondary interface. *Research in Science Education, 9*, 75-84.

POWER, C. (1981). Changes in students' attitudes toward science in the transition between Australian elementary and secondary schools. *Journal of Research in Science Teaching, 18*, 33-39.

RAKOW, S.J. and WALKER, C.L. (1985). The status of Hispanic American students in science: Achievement and exposure. *Science Education, 69*, 557-565.

REITZ, J.G. (1973). The flight from science reconsidered: Career choice of science and engineering in the 1950's and 1960's. *Science Education, 57*, 121-134.

ROGOSA, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J.R. Nesselroade and P.B. Baltes (Eds.), *Longitudinal Research in the Study of Behavior and Development* (pp.263-302). New York: Academic Press.

ROSOW, I. (1978). What is a cohort and why? *Human Development, 21*, 65-75.

ROWELL, J.A. and DAWSON, C.J. (1980). Mountain or mole hill: Can cognitive psychology reduce the dimensions of conceptual problems in classroom practice? *Science Education, 64*, 693-708

ROWELL, J.A. and DAWSON, C.J. (1981). Volume, conservation and instruction: A classroom based Solomon Four Group study of conflict. *Journal of Research in Science Teaching, 18*, 533-546.

RYDER, N.B. (1965). The cohort as a concept in the study of social change. *American Sociological Review, 30*, 843-861.

SCHAIE, K.W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*, 92-107.

SCHAIE, K.W. and HERTZOG, C. (1982). Longitudinal methods. In B.B. Wolman (Ed.), *Handbook of Developmental Psychology* (pp.91-115). Englewood Cliffs, NJ: Prentice Hall.

SCHIBECI, R.A. (1984). Attitudes to science — an update. *Studies in Science Education, 11*, 26-59.

SCOTT, N.C., Jr. (1970). Strategy of inquiry and styles of categorization: A three-year exploratory study. *Journal of Research in Science Teaching, 7*, 95-102.

SCOTT, N.C., Jr. (1973). Cognitive style and inquiry strategies: A five-year study. *Journal of Research in Science Teaching, 10*, 323-330.

SHELDON, D.S. (1978). Long-term impact of curriculum awareness conferences on school administrators and key teachers. *Science Education, 62*, 517-521.

SHIPSTONE, D.M. (1984) A study of children's understanding of electricity in simple DC circuits. *European Journal of Science Education, 6*, 185-198.

SHUELL, T.J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56,* 411-436.

SHULMAN, L.S. and TAMIR, P. (1973). Research on teaching in the natural sciences. In R.M.W. Travers (Ed.), *Second Handbook of Research on Teaching* (pp.1098-1148). Chicago: Rand McNally.

SIMPSON, R.D. and OLIVER, J.S. (1985). Attitude toward science and achievement motivation profiles of male and female science students in grades six through ten. *Science Education, 69,* 511-526.

SMELTZ, J.R. (1966). The study of the retention of learning in high school chemistry for a period of one year. *Science Education, 50,* 359-370.

SNEIDER, C. and PULOS, S. (1983). Children's cosmographies: Understanding the Earth's shape and gravity. *Science Education, 67,* 205-221.

STERRETT, M.D. and DAVIS, R.A. (1954). The permanence of school learning: A review of studies. *Educational Administration and Supervision, 40,* 449-460.

STRAWITZ, B. (1984). Cognitive style and the effects of two instructional treatments on the acquisition and transfer of the ability to control variables: A longitudinal study. *Journal of Research in Science Teaching, 21,* 833-841.

TAMIR, P. and AMIR, R. (1981). Retrospective curriculum evaluation: An approach to the evaluation of long-term effects. *Curriculum Inquiry 11,* 259-278.

TAMIR, P. and JUNGWIRTH, E. (1975). Students' growth as a result of studying BSCS biology for several years. *Journal of Research in Science Teaching, 12,* 263-279.

TOMERA, A.N. (1974). Retention of the science processes of observation and comparison in junior high school students. *Science Education, 58,* 185-194.

TYLER, R.W. (1949). *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press.

WALL, W.D. and WILLMAS, H.L. (1970). *Longitudinal Studies and the Social Sciences.* London: Heinemann.

WALTERS, L.L. (1965). Ninth vs. tenth grade biology — A follow-up study. *Journal of Research in Science Teaching, 3,* 230-234.

WEST, L.H.T. and PINES, A.L. (Eds.). (1985). *Cognitive Structure and Conceptual Change.* Orlando, FL: Academic Press.

WESTERBACK, M.E. (1982). Studies on attitude toward teaching science and anxiety about teaching science in preservice elementary teachers. *Journal of Research in Science Teaching, 19,* 603-616.

WHITE, R.T. (1988). *Learning Science.* Oxford, UK: Blackwell.

WHITE, R.T. and TISHER, R.P. (1986). Research on natural sciences. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 874-905). New York: Macmillan.

WILLSON, V.L. (1980). An investigation of teachers' persistence in implementing NSF-supported science curricula. *Journal of Research in Science Teaching, 17,* 257-261.

WILSON, A.H. and WILSON, J.M. (1984). The development of formal thought during pretertiary science courses in Papua New Guinea. *Journal of Research in Science Teaching, 21,* 527-535.

WOHLWILL, J.F. (1973). *The study of Behavioral Development.* New York: Academic Press.

WRIGHT, E.L. (1978). The influence of intensive instruction upon the open exploration behavior of ninth grade students. *Journal of Research in Science Teaching, 15,* 535-541.

WRIGHT, E.L. (1981). The long-term effect of intensive instruction on the open exploration behavior of ninth grade students. *Journal of Research in Science Teaching, 18,* 557-561.

WYNN, D.C. and BLEDSOE, J.C. (1967). Factors related to gain and loss of scientific interest during high school. *Science Education, 51,* 67-74.